THE CALCULUS OF VARIATIONS.

By Dr. Harris Hancock, Chicago, Ill.

Introduction and General Outline.

Since the time of the Bernouillis, mathematicians have in a greater or less degree considered problems which could be solved by methods of variations. Euler and Lagrange gave these methods more systematic and comprehensive forms, and founded the calculus of variations on a more scientific basis.

By extending these principles other mathematicians have augmented the subject in a wonderful manner; without, however, avoiding many difficulties which arise from want of rigor in the proofs, and from a misinterpretation of some of the fundamental conceptions. All these difficulties were removed when Prof. Karl Weierstrass, in 1879–'80, founded the entire calculus of variations on a new basis, free from any objection, and which at the same time is more comprehensive in its embrace.

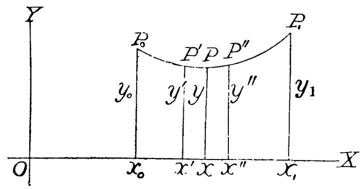
The writer does not think it out of place to bring these investigations before the readers of the Annals of Mathematics, since so little is known of this new treatment of the calculus of variations, especially to American students; and so through the courtesy of the editor he will give separate papers which are, so far as possible, complete in themselves, and which at the same time are intended to include the essential parts of the new theory of variations.

These papers are in a great measure abstracts of lectures that were given at Berlin by Prof. Weierstrass; much is also due to Prof. H. A. Schwarz, whose lectures on the calculus of variations the writer had the pleasure of hearing at Berlin during the summer semester of 1891.

1. *In the differential calculus* a definite function is given, and a special value of the variable or of the variables (if there are more than one variable) is sought, for which the function takes the greatest or the least possible value; in *the calculus of variations* a function is sought, and an expression which depends upon this function in a certain known manner is given. A definite integral is given, in which the integrand depends upon the unknown function in a known manner, and it is asked what form must the unknown function have in order that the definite integral may have a maximum or a minimum value.

We treat only real values of the variables.

2. In order to learn to recognize the general nature of the subject, we shall first state a few problems which may be solved by the calculus of variations; so that, while we seek the general characteristics of these problems, we shall of our own accord come to a more exact statement of the problems which the calculus of variations has to solve.



PROBLEM I. *Two points P_0 and P_1 with coordinates (x_0, y_0) and (x_1, y_1) respectively are given. Both points lie on the same side of the axis of X in the plane xy. It is required to join P_0 and P_1 by a curve such that when the plane is turned through one complete revolution about the axis of x, the zone generated by this curve may have the smallest possible surface.*

3. To show some of the defects of the old methods we proceed as follows: With the assumption that it is possible to draw a curve through the two points which satisfies the conditions of the problem, we suppose that two points $P'(x', y')$ and $P''(x'', y'')$ are taken on the curve, and we find another point $P(x, y)$ on the curve such that

$$x - x' = x'' - x = \Delta x .$$

We suppose that P and P', P and P'' are joined together by straight lines; and later we suppose that these three points are taken very close together, so that there is a transition from the two straight lines to the curve. The remaining portions of the curve on the left hand side of P' and on the right hand side of P'' are supposed to remain unaltered.

The portions of surface generated by the straight lines PP' and PP'' are, respectively,

$$\pi(y' + y)\sqrt{(\Delta x)^2 + (y - y')^2} \text{ and } \pi(y + y'')\sqrt{(\Delta x)^2 + (y'' - y)^2} .$$

The sum of these two surfaces of revolution we consider as a function of the variable y, and it is required to find when

$$\pi(y' + y)\sqrt{(\Delta x)^2 + (y - y')^2} + \pi(y + y'')\sqrt{(\Delta x)^2 + (y'' - y)^2}$$

is a minimum.

In order to have a minimum this expression when differentiated with regard to y must be zero ; i. e.

$$\pi \sqrt{(\Delta x)^2 + (y - y')^2} + \pi \sqrt{(\Delta x)^2 + (y'' - y)^2}$$
$$+ \frac{\pi (y' + y)(y - y')}{\sqrt{(\Delta x)^2 + (y - y')^2}} - \frac{\pi (y + y'')(y'' - y)}{\sqrt{(\Delta x)^2 + (y'' - y)^2}} = 0 . \tag{A}$$

y may be determined from this equation,

$$y = f(x), \text{ say.}$$

Therefore

$$y' = f(x - \Delta x),$$

and

$$y'' = f(x + \Delta x).$$

Hence, by Taylor's theorem,

$$y' = f(x - \Delta x) = f(x) - f'(x)\,\Delta x + \tfrac{1}{2} f''(x)(\Delta x)^2 - \ldots ,$$
$$y'' = f(x + \Delta x) = f(x) + f'(x)\,\Delta x + \tfrac{1}{2} f''(x)(\Delta x)^2 + \ldots .$$

Hence

$$y - y' = f'(x)\,\Delta x - \tfrac{1}{2} f''(x)(\Delta x)^2 + \ldots ,$$
$$y'' - y = f'(x)\,\Delta x + \tfrac{1}{2} f''(x)(\Delta x)^2 + \ldots .$$

Substituting these values in (A), we have

$$\Delta x \sqrt{1 + f'(x)^2 + \ldots} + \Delta x \sqrt{1 + f'(x)^2 + \ldots}$$
$$+ \frac{[2 f(x) - f'(x)\,\Delta x + \ldots][f'(x)\,\Delta x - \tfrac{1}{2} f''(x)\,\Delta x^2 + \ldots]}{\Delta x \sqrt{1 + f'(x)^2}}$$
$$- \frac{[2 f(x) + f'(x)\,\Delta x + \ldots][f'(x)\,\Delta x + \tfrac{1}{2} f''(x)\,\Delta x^2 + \ldots]}{\Delta x \sqrt{1 + f'(x)^2}} = 0 .$$

Dividing through by Δx and making $\Delta x = 0$, we have

$$1 + f'(x)^2 - f(x) f''(x) = 0 ; \tag{B}$$

i. e.

$$1 + \left[\frac{dy}{dx}\right]^2 - y \frac{d^2 y}{dx^2} = 0 .$$

Therefore in order to have a minimum value, $f(x)$ or y must satisfy this differential equation; however, when y satisfies this differential equation, we do not always have a minimum, as will be shown later.

Differentiate (B) with regard to x, and we have

$$\frac{dy}{dx}\cdot\frac{d^2y}{dx^2} = y\,\frac{d^3y}{dx^3},$$

or

$$\frac{\frac{dy}{dx}}{y} = \frac{\frac{d}{dx}\left\{\frac{d^2y}{dx^2}\right\}}{\frac{d^2y}{dx^2}}.$$

That is, $y = c^2\,\frac{d^2y}{dx^2}$, where c^2 is the constant of integration.

Since $y = e^{\frac{x}{c}}$, and $y = e^{-\frac{x}{c}}$ are two solutions of this last differential equation, the most general solution is

$$y = c_1 e^{\frac{x}{\ }} + c_2 e^{-\frac{x}{c}},$$

where c_1 and c_2 are also constants.

This last equation is the equation of the catenary.

4. Thus, by the help of the theory of maxima and minima, we have, it is true, come to a certain result; but, on the other hand, we have yet to ask whether this curve gives a true minimum, and owing to the manner in which we have come to the result, we have yet to see whether this curve only in a definite portion or throughout its whole extent possesses the property required in the problem.

That we are justified in insisting upon this last statement is seen from what follows later, where it will be shown that the curve found above satisfies the required conditions only between given limits.

A simple consideration shows that the method we have followed above is not at all rigorous; since it presupposes, which of itself is not admissible, that the curve which satisfies the problem is regular in its whole extent, since otherwise the portions of curve between the two points $(x - \Delta x, y')$ and (x, y) could not be replaced by straight lines joining these two points.

5. The characteristic difference between problems relative to maxima and minima and the problems which have to do with the calculus of variations, consists in the fact, that in the first case we have to deal with only a *finite number of discrete points*, while in the calculus of variations the question is concerning a *continuous series of points.*

If we wish to substitute in the place of the curve first a polygonal line and afterwards apply to this line methods similar to those used above, then it turns out that, after we have found a line which satisfies all the conditions, it is necessary yet to prove that the required limiting transition from polygonal

line to curve *in reality* results in a definite curve which satisfies the conditions of the problem.

The method given above has been chosen to make clear what is in common between, as well as the difference between, the theory of maxima and minima and the calculus of variations, and we shall now formulate the problem in a different manner.

6. Every limiting transition, as from polygon to curve, is made of itself, if we make use of the conception of integration ; since an integral represents the limiting value of a sum of quantities which increase following a definite law so as to become infinite in number, the quantities themselves becoming smaller and smaller in a corresponding manner.

If we, therefore, define the surface area of the curve $y = f(x)$ which we have to find by

$$S = 2\pi \int y ds ,$$

or

$$\frac{S}{2\pi} = \int_{x_0}^{x_1} y \sqrt{1 + \left[\frac{dy}{dx}\right]^2}\, dx ,$$

then this integral will have a definite value for every curve that is drawn between the points P_0 and P_1, and consequently the problem may be stated as follows :—

PROBLEM I. *y is to be so determined as a function of x, that the above integral shall have the smallest possible value.*

The solution of this problem will be given later.

7. As a second problem may be given the problem of the *brachistochrone* (curve of quickest descent) which may be stated as follows :—

PROBLEM II. *In a vertical plane a curve is to be drawn from a point A to a point B below in such a manner that a material point which is acted upon by gravity, and which is compelled to move upon this curve, shall with a given initial velocity go from A to B in the shortest possible time.*

Let the mass of the material point be 1, its initial velocity a, the acceleration of gravity $2g$, the time t, and the coordinates of A and B respectively (0, 0) and (a, b).

Let the direction of the positive y-axis be the direction of a falling body (gravity), and let the positive x-axis be directed towards the side on which B lies. Then according to the law of the *conservation of energy*,

$$\left[\frac{dx}{dt}\right]^2 + \left[\frac{dy}{dt}\right]^2 = 4gy + a^2$$

or

$$dt = \frac{\sqrt{dx^2 + dy^2}}{\sqrt{4gy + \alpha^2}} = \frac{\sqrt{1 + \left[\frac{dx}{dy}\right]^2}\,dy}{\sqrt{4gy + \alpha^2}};$$

whence

$$T = \int_0^b \frac{\sqrt{1 + \left[\frac{dx}{dy}\right]^2}}{\sqrt{4gy + \alpha^2}}\,dy\,.$$

Our problem then is to so determine x as a function of y that the above integral shall have the smallest possible value.

8. In the two problems given above one of the variables is a one-valued function of the other; this is due to the fact that the system of coordinates may be so chosen in both cases. Since this is not possible in all cases, it is expedient to represent the curve by two equations, that is, to consider x and y as one-valued functions of any quantity t, where t has only the property, that when it goes through all values between two given limits, the corresponding point x, y traverses the curve from the beginning point to the end point, and in such a way that to a greater value of t there corresponds a later* point of the curve. Hence the integrals of our two problems, which are to have a minimum value may be expressed in the form

$$S = 2\pi \int_{t_0}^{t_1} y\sqrt{x'^2 + y'^2}\,dt\,, \tag{I}$$

$$T = \int_{t_0}^{t_1} \frac{\sqrt{x'^2 + y'^2}}{\sqrt{4gy + \alpha^2}}\,dt\,, \tag{II}$$

where

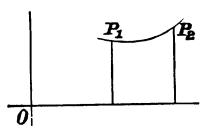
$$x' = \frac{dx}{dt}\,, \quad y' = \frac{dy}{dt}\,.$$

9. Problem III. *Between two points on a regular surface $f(x, y, z) = 0$, a curve is to be drawn so that its length is a minimum.*

Consider the orthogonal coordinates x, y, z of a point of the surface represented as one-valued regular functions of two parameters u, v. If we consider these as the rectangular coordinates of a point of the plane, then to every

---

* If $t_1 < t_2$ and the point P_1 corresponds to t_1, and P_2 to t_2, then P_2 in reference to P_1 is known as a *later* point, and P_1 in reference to P_2 as an *earlier* point.



point of the surface there will correspond a definite point of the uv-plane, and these points in their collectivity fill out a definite part of the plane, which may be looked upon as the image of the surface on the plane. To every curve on the surface corresponds a curve in this part of the uv-plane and reciprocally.

Consider, further, u and v as onevalued functions of a quantity t; hence to every value t there corresponds a point of the uv-plane, and therefore, also, in case this point lies in the definite part of the uv-plane, there is a corresponding definite point of the surface.

Consequently if t_0 and t_1 are values of t which correspond to the two fixed points on the surface, then the length of any curve which lies between these two points is determined through

$$L = \int_{t_0}^{t_1} \sqrt{P\left[\frac{du}{dt}\right]^2 + 2Q\frac{du}{dt}\cdot\frac{dv}{dt} + R\left[\frac{dv}{dt}\right]^2}\, dt,$$

where

$$P = \left[\frac{\partial x}{\partial u}\right]^2 + \left[\frac{\partial y}{\partial u}\right]^2 + \left[\frac{\partial z}{\partial u}\right]^2,$$

$$Q = \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v},$$

$$R = \left[\frac{\partial x}{\partial v}\right]^2 + \left[\frac{\partial y}{\partial v}\right]^2 + \left[\frac{\partial z}{\partial v}\right]^2.$$

We have to determine u and v as functions of t, so that L is a minimum.

10. What is common to these three problems is that we have to determine two functions of t in such a way that an integral depending upon them and of the form

$$I = \int_{t_0}^{t_1} F(x, y, x', y')\, dt$$

may have the smallest possible value.

Here t_0 and t_1 have fixed values, and $F(x, y, x', y')$ represents a function of x, y, x', y', of which x', y' may be regarded as unlimited variables, while x, y are limited to a region which extends over the whole plane, or over a continuous part of the same.

11. The condition that t_0, t_1 should have fixed values is not essential; moreover both end points may move, as in the case of the third problem, if we give it the following form: Two curves are given on a surface; among all the possible curves between the points of the one curve and the points of the other curve, that curve is to be found which has the shortest length. We are accustomed to call this the geodesic distance of the two curves.

In order to solve this problem, we must first solve the special problem III, since if a curve has the property of being of minimum length such as is required above, it must also retain the same property, if we consider the end points fixed. Hence, from III the nature of the curve must be determined. The variation of the end points give in addition certain special properties, which the curve must possess.

For example, the shortest distance between two curves which lie in the same plane is clearly a straight line; through the variation of the end points it follows that this straight line must be at the same time perpendicular to both curves.

12. PROBLEM IV. Essentially different from the three problems given above is the following: *We are so to construct a closed curve with given periphery, that the surface inclosed shall have the greatest possible area.*

Consider x and y such functions of t, say $x(t)$, $y(t)$, that for two definite values t_0 and t_1 of t, the corresponding points x, y of the curve fall together, and that, if t goes from a smaller value t_0 to a greater value t_1, x, y completely traverses the curve in the positive direction; then twice the area of the surface inclosed by the curve will be expressed by the integral

$$I = \int_{t_0}^{t_1} (xy' - yx')\, dt\,,$$

and the periphery of the curve is determined by means of

$$I_1 = \int_{t_0}^{t_1} \sqrt{x'^2 + y'^2}\, dt\,.$$

13. The proposed problem is now as follows: *So determine x, y as functions of t, that the integral I which depends upon them, shall have the greatest possible value, while at the same time I_1 has a given fixed value.*

Problems of this nature are the most interesting and of the most frequent occurrence. They require a treatment essentially different from that of those first mentioned.

These problems are sufficient to give a conception of the nature of the problems which are to be solved by the calculus of variations, and with these as a basis it will be possible to define the object of the calculus of variations.

We must yet, however, introduce the fundamental conception of the variation of a curve. In former times the calculus of variations was considered one of the most difficult branches of analysis; it was wrongly thought that the ground of this difficulty was in the supposed lack of clearness in the fundamental conceptions, especially in the conception of the variation of a curve,

while the difficulties that do arise, lie for the most part in quite another direction.

14. In the theory of maxima and minima we say the value of a function is, for a definite system of values of the variables, a maximum or a minimum, if this value of the function for this system of values is greater or smaller than for all the neighboring systems of values.

We say of a function $f(x)$ of one variable, it has at a definite position $x = a$, a maximum or a minimum value, if the value for $x = a$ is respectively greater or less than it is for all other values of x which are situated in the neighborhood $|x - a| < \delta$ as near as we wish to a.

The analytical condition that $f(x)$ shall have for the position $x = a$

$$\left.\begin{array}{lll} \textit{a maximum} \text{ is expressed by } & f(x) - f(a) < 0\,; \\ \textit{a minimum} \quad \text{“} \qquad \text{“} & f(x) - f(a) > 0\,. \end{array}\right\} |x - a| < \delta.$$

In the same way, *we say of a function $f(x_1, x_2, \ldots, x_n)$ of n variables, that it has for a definite position $x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n$, a maximum or a minimum, if the value of the function for $x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n$ is respectively greater or smaller than it is for all other systems of values, which are situated in the neighborhood $|x_\lambda - a_\lambda| < \delta_\lambda$ ($\lambda = 1, 2, \ldots, n$), as near as we wish to the first position.*

As here we speak of a neighboring system of values, so also in the calculus of variations, we speak of curves which lie in the neighborhood of a given curve; and we require that an integral in the case of a minimum should be less and in the case of a maximum greater when taken over the given curve than for any of the neighboring curves.

15. In order to fix the conception of a neighboring curve, and to make clear the analogy of the same with the conception of a neighboring system of values, let us consider first instead of the given curve a broken line $A_1A_2A_3 \ldots A_n$, and let us cause the same to slide just a little from its original position.

Then in the new position every corner B_k will correspond to a definite corner A_k in the old position, and moreover the new position $B_1B_2B_3 \ldots B_n$ will be as little different from the old position $A_1A_2A_3 \ldots A_n$ as we wish, if we stipulate that the distance between any two corresponding points A_k and B_k shall be smaller than any quantity δ where δ is as small as we choose. Now, by increasing the number of sides, let the broken line pass into the given curve, then the points $B_1B_2 \ldots B_n$ will also form a curve which is little different from the first curve, and which we consequently call *neighboring to* the first curve.

We can, therefore, say a curve *is neighboring to* another curve, or exists

out of another curve through a variation as small as we choose, if to every point of the latter curve there corresponds a definite point on the former curve and also the distance between any two corresponding points is smaller than δ, where δ is as small as we choose.

This geometrical conception of a neighboring curve offers no obscurity. In a similar manner it is easy to see that for every change of the curve, there is a corresponding change of the integral

$$\int F(x, y, x', y')\, dt\,,$$

and that this change will be infinitely small, when the second curve is neighboring to the first.

This change of the value of the integral must of course be a continuous negative one, if the integral is to be a maximum, and a continuous positive one, if the integral is to be a minimum.

16. In accordance with this we may formulate the problem of the calculus of variations as follows :

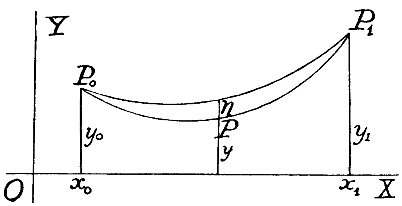
The quantities x, y are to be so determined as functions of a quantity t, that when we define a curve by the equations $x = x(t)$, $y = y(t)$, and cause the curve to vary as little as we choose, the change which in consequence takes place in the integral

$$I = \int_{t_0}^{t_1} F(x, y, x', y')\, dt$$

is continuously positive if a minimum, and continuously negative if a maximum is to enter.

If we consider the problem as proposed in this manner, we have a definite problem of the calculus of variations before us, and we have to find strict and rigorous methods for the solution of this problem.

17. *Variation of curves.*



Let us return for a moment to the integral

$$S = \int_{x_0}^{x_1} 2\pi y \sqrt{1 + \left[\frac{dy}{dx}\right]^2}\, dx\,. \tag{1}$$

Let $y = f(x)$ be the curve which gives a minimum surface area when we rotate this curve about the axis of x.

Let η be the distance between this curve and any neighboring curve measured on the y-ordinate, and suppose that η is a continuous function of x subject to the conditions, that for $x = x_0$, $\eta = 0$; for $x = x_1$, $\eta = 0$; and for all other points $|\eta| < \rho$, where ρ may be as small as we choose.

$$\eta' = \frac{d\eta}{dx} \text{ and } \int_{x_0}^{x_1} \eta' dx = [\eta]_{x_0}^{x_1} = 0 .$$

The integral of any neighboring curve corresponding to (1) is

$$\int_{x_0}^{x_1} 2\pi (y + \eta) \sqrt{1 + \left[\frac{d(y + \eta)}{dx}\right]^2} dx . \tag{2}$$

Hence the total variation caused in (1) when, instead of $y = f(x)$, we take a neighboring curve, is

$$\Delta S = \int_{x_0}^{x_1} 2\pi (y + \eta) \sqrt{1 + \left[\frac{d(y + \eta)}{dx}\right]^2} dx - \int_{x_0}^{x_1} 2\pi y \sqrt{1 + \left\{\frac{dy}{dx}\right\}^2} dx . \tag{3}$$

ΔS has always a positive sign since the surface in question is a minimum.

18. Instead of the one neighboring curve, we may consider a whole bundle of such curves, if for η we substitute $\varepsilon\eta$, where ε is independent of x and has any value between -1 and $+1$. After this substitution (3) becomes

$$\Delta S = \pi \left[\int_{x_0}^{x_1} 2(y + \varepsilon\eta) \sqrt{1 + \left[\frac{d}{dx}(y + \varepsilon\eta)\right]^2} dx - \int_{x_0}^{x_1} 2y \sqrt{1 + \left\{\frac{dy}{dx}\right\}^2} dx\right]; \tag{4}$$

and, on the other hand, developing ΔS by Taylor's theorem,

$$\Delta S = \varepsilon \delta S + \frac{\varepsilon^2}{1 \,.\, 2} \delta^2 S + \frac{\varepsilon^3}{1 \,.\, 2 \,.\, 3} \delta^3 S + \ldots . \tag{5}$$

There is no constant term in this last development, since when ε is made zero in (4) the first and second integrals cancel each other.

δS is the first variation,

$\delta^2 S$ is the second variation, etc.

Instead of taking η a very small quantity, we may take ε so small that $\varepsilon\eta$ is as small as we choose.

With Lagrange, writing $\eta = \delta y$, it is seen that the total change in y is $\epsilon\eta = \epsilon\delta y = \Delta y$.

REMARK. The sign of differentiation and the sign of variation may be interchanged; for example the 1st derivative of a variation is equal to the 1st variation of a derivative, as is seen by writing

$$\eta = \delta y\,, \text{ then } \eta' = (\delta y)' = \frac{d}{dx}(\delta y)\,. \tag{1}$$

Again $\eta = \delta y$; change y into $y + \epsilon\eta$, and consequently y' into $y' + \epsilon\eta'$; whence

$$\eta' = \delta y' = \delta\left[\frac{dy}{dx}\right]; \tag{2}$$

hence, from (1) and (2),

$$\frac{d}{dx}(\delta y) = \delta\left[\frac{dy}{dx}\right].$$

19. Returning to (4) and writing $y' = \dfrac{dy}{dx}$, $\eta' = \dfrac{d\eta}{dx}$ and expanding the expression under the sign of integration

$$2\pi(y + \epsilon\eta)\sqrt{1 + (y' + \epsilon\eta')^2} - 2\pi y\sqrt{1 + y'^2}\,,$$

we have

$$\pi\epsilon\left[2\eta\sqrt{1 + (y' + \epsilon\eta')^2} + \frac{(y + \epsilon\eta)(y' + \epsilon\eta')\eta'}{\sqrt{1 + (y' + \epsilon\eta')^2}}\right]_{\epsilon = 0} + \epsilon^2(\ldots)\,.$$

Hence, equating the coefficients of the 1st power of ϵ in (4) and in (5) we have

$$\delta S = 2\pi\int_{x_0}^{x_1}\left[\sqrt{1 + y'^2}\,.\,\eta + \frac{yy'}{\sqrt{1 + y'^2}}\eta'\right]dx\,,$$

which is a homogeneous function of the 1st degree in η and η' (η' cannot be infinitely large, since then the development would not be necessarily convergent).

In a similar manner we may find a definite integral for the 2nd variation, in which the integrand is an integral homogeneous function of the 2nd degree in η and η'; similarly for the third variation, etc.